MULTIVARIATE REGIONAL DEEP LEARNING PREDICTION OF SOIL PROPERTIES FROM NEAR-INFRARED, MID-INFRARED AND THEIR COMBINED SPECTRA

#11647

[1,2*]Rumbidzai W. Nyawasha, [3]Alexandre M.J.-C. Wadoux, [4,5]Pierre Todoroff, [1,6]Regis Chikowo, [1,2,5,7]Gatien N. Falconnier, [4,5]Maeva Lagorsse, [5,8]Marc Corbeels, [1,2,5]Rémi Cardinael
[1]Plant Production Sciences and Technology, University of Zimbabwe, Harare, Zimbabwe
[2]CIRAD, UPR AIDA, Harare, Zimbabwe
[3]LISAH, Université de Montpellier, INRAE, IRD, Institut Agro, Montpellier, France
[4]CIRAD, UPR AIDA, F-97410 Saint-Pierre, Réunion, France
[5]AIDA, Univ Montpellier, CIRAD, Montpellier, France
[6]Plant, Soil and Microbial Sciences Department, Michigan State University, East Lansing, MI 48824, USA
[7]International Maize and Wheat Improvement Centre (CIMMYT)-Zimbabwe, Harare, Zimbabwe
[8]IITA, International Institute of Tropical Agriculture, PO Box 30772, Nairobi 00100, Kenya
[*]e-mail: rumbidzaiwnyawasha@yahoo.com; +263 776 163 414

ABSTRACT

Artificial neural network (ANN) models have been successfully used in infrared spectroscopy research for the prediction of soil properties. They often show better performance than conventional methods such as partial least squares regression (PLSR). In this paper we develop and evaluate a multivariate extension of ANN for predicting correlated soil properties: total carbon (C), total nitrogen (N), clay, silt, and sand contents, using visible near-infrared (vis-NIR), mid-infrared (MIR) or combined spectra (vis-NIR + MIR). We hypothesize that accounting for the correlation through joint modelling of soil properties with a single model can eliminate "pedological chimera": unrealistic values that may arise when properties are predicted independently such as when calculating ratio or soil texture values. We tested two types of ANN models, a univariate (ANN-UV) and a multivariate model (ANN-MV), using a dataset of 228 soil samples collected from Murehwa district in Zimbabwe at two soil depth intervals (0 – 20 and 20 - 40 cm). The models were compared with results from a univariate PLSR (PLSR-UV) model. We found that the multivariate ANN model was better at conserving the observed correlations between properties and consequently gave realistic soil C:N and C:Clay ratios, but that there was no improvement in prediction accuracy over using a univariate model (ANN or PLSR). The use of combined spectra (vis-NIR + MIR) did not make any significant improvements in prediction accuracy of the multivariate ANN model compared to using the vis-NIR or MIR only. We conclude that the multivariate ANN model is better suited for the prediction of multiple correlated soil properties and that it is flexible and can account for compositional constrains. The multivariate ANN model helps to keep realistic ratio values – with strong implications for assessment studies that make use of such predicted soil values.

INTRODUCTION

Soils and soil properties vary over space in relation to the parent material, climate, topography, among others, and change over time in response to natural processes and human activities (Jenny, 1994; Beillouin et al., 2023). Sampling and monitoring of soils is costly and time consuming, as it usually requires a large number of measurements and laboratory analyses (Webster and Lark, 2013). To adequately capture the spatial and temporal variations of soils, effective and less costly methods of data collection and analysis have been developed, including the use of visible and near-infrared (vis-NIR) and mid-infrared (MIR) spectroscopy (Nocita et al., 2015). Statistical models can then be employed to establish a predictive relationship between the spectral characteristics and values of soil properties for which corresponding laboratory measurements are available (Barra et al., 2021). Partial least squares regression (PLSR) has become the most popular regression model in soil spectroscopy (Viscarra Rossel and Lark, 2009; Soriano-Disla et al., 2014). It has been shown to perform well in many situations (Janik et al., 1998; Viscarra Rossel et al., 2006; Cambou et al., 2016; Allo et al., 2020; Bachion de Santana and Daly, 2022). Usually, each soil property is modelled independently, ignoring the correlations that exist between properties. In cases where multiple dependent properties are predicted, this can result in inconsistent predictions and the occurrence of "pedological chimera" as defined by Lagacherie et al. (2022). As a solution, multivariate counterparts of PLSR have been developed, the most common being the PLS2 regression model, a modification of PLSR developed by Wold et al. (1983) and Martens and Naes (1987). However, in terms of predictive accuracy, PLS2 usually performs worse than a model fitted for an individual variable. Several studies, (Pedro and Ferreira, 2007; Blanco and Peguero, 2008; Mishra and Passos, 2022), acknowledged that the univariate model gave higher prediction accuracy than PLS2.

Recently, data-driven models and algorithmic tools from the field of machine learning have become popular for predicting soil properties from spectral data (Meza Ramirez et al., 2021). Commonly used algorithms in soil spectroscopy are support vector machines (Demattê and da Silva Terra, 2014; Deiss et al., 2020), cubist (Minasny and McBratney, 2008; Clergue et al., 2023), random forest (Viscarra Rossel and Behrens, 2010; McDowell et al., 2012; Wadoux, 2023), and artificial neural networks (ANNs) (Daniel et al., 2003; Wijewardane et al., 2018). The use of ANNs has been successful for soil property prediction and showed better performance than conventional methods such as PLSR in several studies (Daniel et al., 2003; Viscarra Rossel and Behrens, 2010; Ng et al., 2019; Padarian et al., 2019). The main advantages of ANNs over conventional regression models are the ability to extract relevant information in high-dimensional datasets, the modelling of non-linear relationships between spectra and soil properties, and a flexibility in the definition of the algorithm and objective function (Ludwig et al., 2019; Margenot et al., 2020). Despite its flexibility, to date very few studies have attempted to understand whether a multivariate ANN model accounts for the correlations that exist amongst soil properties, although promising results were found in Mishra and Passos (2022), Ng et al. (2019), and Ramsundar et al. (2015).

In this paper we develop, further expand, and test the multivariate extension of ANNs for predicting soil properties from their vis-NIR, MIR and combined spectra (vis-NIR + MIR). After model development, we investigate the ability of the multivariate model to predict correlated soil properties, as compared to a model that predicts each property individually. The methodology is tested for total carbon, total nitrogen, sand, silt, and clay contents in soils from Murehwa district located in the sub-humid region of Zimbabwe. We hypothesize that combined modelling of several soil properties can eliminate "pedological chimera" by accounting for the correlations between the properties. The comparison between observed and predicted soil properties from a univariate and a multivariate model is made using vis-NIR, MIR or combined vis-NIR + MIR spectra.

METHODOLOGY

The study was done in Murehwa district (17°39'S, 31°47'E), a smallholder farming area situated about 80 km northeast of Harare, the capital city of Zimbabwe. Soil samples were collected in three villages randomly selected from Ward 28 of the district. 50 % of the households in the three villages were then randomly selected to give a total of 183 farming households. Soil samples were collected from all agricultural fields belonging to the selected households. Samples were also collected from common lands – lands that are available for all villagers and used for grazing, collecting firewood, litter, and wild fruits. Soil samples were collected between June and July 2021 at two depths i) 0 – 20 cm ii) 20 – 40 cm. Sampling was carried out following a zig-zig transect covering each field, with a sub-sample being collected at 10 m distance using an auger and all the sub-samples were mixed to obtain a composite per field and depth.

Spectra were acquired at the laboratory of the French Agricultural Research Centre for International Development (CIRAD) in Saint Denis, La Réunion, on all soil samples ground to 200µm. The MIR spectra were measured using an Agilent 4300 handheld FTIR spectrometer (Agilent Technologies, Santa Clara, CA) over a wavenumber range between 650 – 4000 cm^{-1} with a measurement interval of 4 cm^{-1}, vis-NIR spectra were measured using a LabSpec 5000 (Analytical Spectral Devices, Inc. Boulder, CO, USA) with an optical fibre connected to the internal light (adapted to small sample sizes) over a wavelength of 350 – 2500nm and spectral resolution of 3 nm at 700 nm and 10 nm at 1400/2100 nm. Spectral pre-processing was done to ensure the removal of any variations caused by light scattering and to enhance some features within the spectra (Wadoux et al., 2021). The MIR spectra were trimmed to remove the noise at the edges leaving the range between 800-4000 cm^{-1} whereas vis-NIR spectra were trimmed to 20000 – 4080 cm^{-1}. The MIR and vis-NIR datasets were then combined using spectra concatenation to create a third dataset (vis-NIR + MIR) ranging between 10000 - 800 cm^{-1}. A For laboratory analysis, a subset of 230 soil samples, corresponding to 17 % of the total number of samples, was selected. The selection was based on spectra similarity and the most representative spectra were chosen using the Kennard Stone algorithm as implemented in the Unscrambler X 10.5 Software (CAMO Software Inc., Oslo, Norway). Total carbon and total nitrogen were determined by the Dumas elemental dry combustion method using an Elementar VarioMax Cube. Soil texture analysis was done using the hydrometer method following Gee and Bauder (1986).

Two types of ANN models were built, a univariate model which predicts one soil property at a time, and a multivariate model which predicts more than one property at the same time. The univariate model was made up of one input layer, three hidden layers and one output layer. The multivariate model was made up of one input layer, four hidden layers and an output layer predicting five outputs simultaneously. The models were trained using vis-NIR, MIR and the combined vis-NIR + MIR data. The two ANN models were compared to a univariate PLSR model to gauge their performance against a conventional model. The ANN models in this study were built using the *keras* package (Allaire and Chollet, 2023) in R with *tensorflow* as backend (Allaire and Tang, 2023) and the PLSR was built using the *pls* package (Liland et al., 2023) also in R. The measured values of the soil properties from the laboratory analyses were used to fit the models. The measured values were split into training and validation sets using k-fold cross-validation to assess prediction accuracy of the model predictions on unseen data. Validation statistics – i.e. mean error (ME), the root mean square error (RMSE) and the coefficient of determination R^2 - were calculated from the pairwise comparison of measured and predicted values obtained from all folds as each represents a specific aspect of prediction quality.

RESULTS AND DISCUSSION

The best prediction models were obtained using MIR spectra, followed by vis-NIR + MIR spectra and lastly by vis-NIR spectra. Model predictions based on MIR spectra had consistently higher R^2 values and lower RMSE values, and this difference was significant when compared to predictions based on vis-NIR spectra (Table 1). This can be attributed to the presence of fundamental vibrations in the MIR region whereas only overtones and combinations bands are present in the vis-NIR regions. Other studies report similar results, particularly for soil carbon predictions where MIR outperforms vis-NIR (Viscarra Rossel et al., 2006; Vohland et al., 2014; Wijewardane et al., 2018). The use of combined vis-NIR + MIR spectra did not improve the predictive accuracy of soil properties in this study. There are varying results on this - a study conducted by Johnson et al. (2019) reported an improved accuracy with combined spectra for several soil properties whereas others report that because the predictions with MIR spectra alone are already highly accurate, combining spectra either results in slightly worse results (Viscarra Rossel et al., 2006; Shao and He, 2011; Ng et al., 2019) or produces results that are equally comparable to MIR alone (Knox et al., 2015).

Table 1. Comparison of the PLSR-UV, ANN-UV and ANN-MV models for three spectral datasets, vis-NIR, MIR and combined vis-NIR + MIR using mean error (ME), root mean square error (RMSE) and coefficient of determination (R^2)

	vis-NIR				MIR				vis-NIR + MIR			
	Model	ME	RMSE	R^2	Model	ME	RMSE	R^2	Model	ME	RMSE	R^2
Total C	PLSR-UV	0.07	4.78	0.74	PLSR-UV	0.09	2.87	0.91	PLSR-UV	0.05	3.99	0.82
	ANN-UV	−0.24	5.41	0.66	ANN-UV	0.52	3.13	0.89	ANN-UV	0.06	4.99	0.71
	ANN-MV	−2.69	6.66	0.49	ANN-MV	−0.79	3.09	0.89	ANN-MV	−1.13	4.23	0.79
Total N	PLSR-UV	0.00	0.35	0.72	PLSR-UV	0.01	0.24	0.87	PLSR-UV	0.01	0.31	0.78
	ANN-UV	0.05	0.43	0.59	ANN-UV	0.01	0.28	0.83	ANN-UV	0.02	0.41	0.64
	ANN-MV	−0.18	0.48	0.48	ANN-MV	−0.09	0.26	0.85	ANN-MV	−0.07	0.31	0.78
Sand	PLSR-UV	0.12	9.19	0.63	PLSR-UV	0.19	6.68	0.80	PLSR-UV	0.05	7.08	0.78
	ANN-UV	−2.86	10.38	0.52	ANN-UV	−1.04	7.28	0.77	ANN-UV	−0.48	9.38	0.61
	ANN-MV	2.63	10.2	0.54	ANN-MV	0.17	7.15	0.77	ANN-MV	1.39	7.72	0.74
Clay	PLSR-UV	−0.07	7.15	0.59	PLSR-UV	−0.07	5.68	0.74	PLSR-UV	−0.03	5.82	0.73
	ANN-UV	−0.13	7.53	0.55	ANN-UV	−0.65	6.18	0.69	ANN-UV	−0.98	6.62	0.65
	ANN-MV	−1.75	7.83	0.51	ANN-MV	−0.08	5.81	0.73	ANN-MV	0.98	5.92	0.72
Silt	PLSR-UV	−0.06	4.15	0.36	PLSR-UV	−0.02	3.49	0.55	PLSR-UV	0.05	3.67	0.50
	ANN-UV	−0.27	5.29	0.68	ANN-UV	−0.56	3.56	0.53	ANN-UV	0.06	4.99	0.71
	ANN-MV	−0.83	4.31	0.31	ANN-MV	−0.04	3.42	0.57	ANN-MV	−0.35	3.71	0.49

We also studied the predictions of two key ratios: *(1)* soil C:N ratio, which is calculated using total carbon and total nitrogen values and is a sensitive indicator of soil quality and for assessing the carbon and nitrogen nutrition balance of soils; and *(2)* the C:Clay ratio, calculated using soil carbon and clay content, which has been proposed as an indicator for soil organic carbon status and soil structure quality (Poeplau and Don, 2023). The range of values for the soil C:N ratio was all within the range between 10 – 25, comparable to the measured values, whereas the ANN-UV model gave more unrealistic values including some negative ones. Previous studies in the study area have shown that soil carbon concentrations in the most fertile soils rarely exceed 10 g C kg^{-1} (Masvaya et al., 2010; Zingore et al., 2011). For the C:Clay ratios the predictions made by the ANN-MV model gave significantly better results (Figure 2). Soil clay content plays an important role in the stabilization of SOC since clay minerals have a high specific surface area and carry a charge, enabling them to bind, and thereby chemically stabilize, organic matter. Clay aggregates also provide micropores for the physical protection of soil organic carbon (Wattel-Koekkoek et al., 2001). The C:Clay ratios obtained in this study range between 1:10 – 1:13 and sometimes even lower, which suggests that these soils are degraded (Poeplau and Don, 2023). This is accurate as these soils are granitic derived. A low clay plus silt fraction usually provides little physical protection of organic matter to influence soil physical properties (Feller and Beare, 1997; Nyamangara et al., 2014). Moreover, clay content is not an accurate predictor of SOC, particularly in tropical soils with high concentrations of aluminium and iron oxides (Khomo et al., 2017; Kirsten et al., 2021).

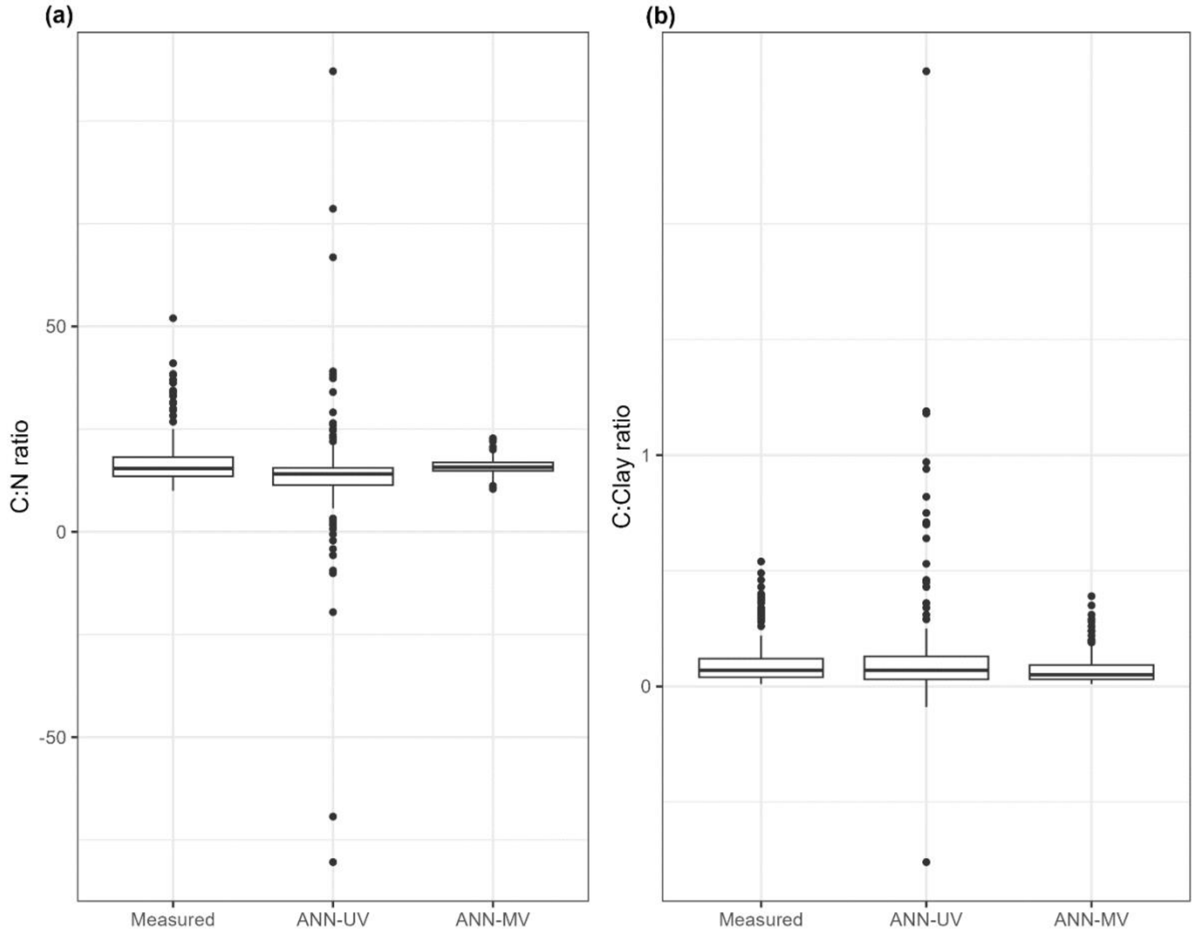

Figure 1. Boxplots of a) soil C:N ratio and b) C:Clay ratio calculated with measured values and the two ANN models

REFERENCES

Allaire, J.., Chollet, F., 2023. Keras: R interface to 'Keras' [WWW Document]. URL https://github.com/rstudio/keras (accessed 8.23.23).

Allaire, J., Tang, Y., 2023. Tensorflow: R Interface to 'TensorFlow' [WWW Document]. R Packag. version 2.11.0.9000. URL https://github.com/rstudio/tensorflow%0A (accessed 8.23.23).

Allo, M., Todoroff, P., Jameux, M., Stern, M., Paulin, L., Albrecht, A., 2020. Prediction of tropical volcanic soil organic carbon stocks by visible-near- and mid-infrared spectroscopy. Catena 189, 104452. https://doi.org/10.1016/j.catena.2020.104452

Bachion de Santana, F., Daly, K., 2022. A comparative study of MIR and NIR spectral models using ball-milled and sieved soil for the prediction of a range soil physical and chemical parameters. Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 279, 121441. https://doi.org/10.1016/j.saa.2022.121441

Barra, I., Haefele, S.M., Sakrabani, R., Kebede, F., 2021. Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances–A review. TrAC - Trends Anal. Chem. 135, 116166. https://doi.org/10.1016/j.trac.2020.116166

Beillouin, D., Corbeels, M., Demenois, J., Berre, D., Boyer, A., Fallot, A., Feder, F., Cardinael, R., 2023. A global meta-analysis of soil organic carbon in the Anthropocene. Nat. Commun. 14, 1–10. https://doi.org/10.1038/s41467-023-39338-z

Blanco, M., Peguero, A., 2008. An expeditious method for determining particle size distribution by near infrared spectroscopy: Comparison of PLS2 and ANN models. Talanta 77, 647–651. https://doi.org/10.1016/j.talanta.2008.07.015

Cambou, A., Cardinael, R., Kouakoua, E., Villeneuve, M., Durand, C., Barthès, B.G., 2016. Prediction of soil organic carbon stock using visible and near infrared reflectance spectroscopy (VNIRS) in the field. Geoderma 261, 151–159. https://doi.org/10.1016/j.geoderma.2015.07.007

Clergue, T.C., Saby, N.P.A., Wadoux, A.M.J.-C., Barthès, B.G., Lacoste, M., 2023. Estimating soil aggregate stability with infrared spectroscopy and pedotransfer functions. Soil Secur. 11, 100088. https://doi.org/10.1016/j.soisec.2023.100088

Daniel, K.W., Tripathi, N.K., Honda, K., 2003. Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand). Aust. J. Soil Res. 41, 47–59. https://doi.org/10.1071/SR02027

Deiss, L., Margenot, A.J., Culman, S.W., Demyan, M.S., 2020. Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. Geoderma 365, 114227. https://doi.org/10.1016/j.geoderma.2020.114227

Demattê, J.A.M., da Silva Terra, F., 2014. Spectral pedology: A new perspective on evaluation of soils along pedogenetic alterations. Geoderma 217–218, 190–200. https://doi.org/10.1016/j.geoderma.2013.11.012

Feller, C., Beare, M.H., 1997. Physical control of soil organic matter dynamics in the tropics. Geoderma 79, 69–116. https://doi.org/10.1016/S0016-7061(97)00039-6

Gee, G.W., Bauder, J.W., 1986. Particle-size Analysis, in: Klute, A. (Ed.), Methods of Soil Analysis, Part 1. Physical and Mineralogical Methods-Agronomy. Agronomy Society of America/Soil Science Society of America, Madison, Wisconsin, pp. 384–411.

Janik, L.J., Merry, R.H., Skjemstad, J.O., 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? Aust. J. Exp. Agric. 38, 681–696. https://doi.org/10.1071/EA97144

Jenny, H., 1994. Factors of Soil Formation A System of Quantitative Pedology. Courier Corporation.

Johnson, J.M., Vandamme, E., Senthilkumar, K., Sila, A., Shepherd, K.D., Saito, K., 2019. Near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for assessing soil fertility in rice fields in sub-Saharan Africa. Geoderma 354, 113840. https://doi.org/10.1016/j.geoderma.2019.06.043

Khomo, L., Trumbore, S., Bern, C.R., Chadwick, O.A., 2017. Timescales of carbon turnover in soils with mixed crystalline mineralogies. Soil 3, 17–30. https://doi.org/10.5194/soil-3-17-2017

Kirsten, M., Mikutta, R., Vogel, C., Thompson, A., Mueller, C.W., Kimaro, D.N., Bergsma, H.L.T., Feger, K.H., Kalbitz, K., 2021. Iron oxides and aluminous clays selectively control soil carbon storage and stability in the humid tropics. Sci. Rep. 11, 1–12. https://doi.org/10.1038/s41598-021-84777-7

Knox, N.M., Grunwald, S., McDowell, M.L., Bruland, G.L., Myers, D.B., Harris, W.G., 2015. Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. Geoderma 239–240, 229–239.

https://doi.org/10.1016/j.geoderma.2014.10.019
Lagacherie, P., Buis, S., Constantin, J., Dharumarajan, S., Ruiz, L., Sekhar, M., 2022. Evaluating the impact of using digital soil mapping products as input for spatializing a crop model: The case of drainage and maize yield simulated by STICS in the Berambadi catchment (India). Geoderma 406, 115503. https://doi.org/https://doi.org/10.1016/j.geoderma.2021.115503.
Liland, K.., Mevik, B.-H., Wehrens, R., Hiemstra, P., 2023. pls: Partial Least Squares and Principal Component Regression [WWW Document]. 2.8-2. URL https://github.com/khliland/pls (accessed 8.23.23).
Ludwig, B., Murugan, R., Parama, V.R.R., Vohland, M., 2019. Accuracy of Estimating Soil Properties with Mid-Infrared Spectroscopy: Implications of Different Chemometric Approaches and Software Packages Related to Calibration Sample Size. Soil Sci. Soc. Am. J. 83, 1542–1552. https://doi.org/10.2136/sssaj2018.11.0413
Margenot, A., O' Neill, T., Sommer, R., Akella, V., 2020. Predicting soil permanganate oxidizable carbon (POXC) by coupling DRIFT spectroscopy and artificial neural networks (ANN). Comput. Electron. Agric. 168, 105098. https://doi.org/10.1016/j.compag.2019.105098
Martens, H., Naes, T., 1987. Multivariate calibration by data compression, in: Williams, P.., Norris, K. (Eds.), Near-Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists, St. Paul, MN, pp. 57–87.
Masvaya, E.N., Nyamangara, J., Nyawasha, R.W., Zingore, S., Delve, R.J., Giller, K.E., 2010. Effect of farmer management strategies on spatial variability of soil fertility and crop nutrient uptake in contrasting agro-ecological zones in Zimbabwe. Nutr. Cycl. Agroecosystems 88, 111–120. https://doi.org/10.1007/s10705-009-9262-y
McDowell, M.L., Bruland, G.L., Deenik, J.L., Grunwald, S., Knox, N.M., 2012. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. Geoderma 189–190, 312–320. https://doi.org/10.1016/j.geoderma.2012.06.009
Meza Ramirez, C.A., Greenop, M., Ashton, L., Rehman, I. ur, 2021. Applications of machine learning in spectroscopy. Appl. Spectrosc. Rev. 56, 733–763. https://doi.org/10.1080/05704928.2020.1859525
Minasny, B., McBratney, A.B., 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemom. Intell. Lab. Syst. 94, 72–79. https://doi.org/10.1016/j.chemolab.2008.06.003
Mishra, P., Passos, D., 2022. Multi-output 1-dimensional convolutional neural networks for simultaneous prediction of different traits of fruit based on near-infrared spectroscopy. Postharvest Biol. Technol. 183, 111741. https://doi.org/10.1016/j.postharvbio.2021.111741
Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., McBratney, A.B., 2019. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. Geoderma 352, 251–267. https://doi.org/10.1016/j.geoderma.2019.06.016
Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Dor, E. Ben, Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J.A.M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J.M., Shepherd, K.D., Stenberg, B., Towett, E.K., Vargas, R., Wetterlind, J., 2015. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. Adv. Agron. 132, 139–159. https://doi.org/10.1016/bs.agron.2015.02.002
Nyamangara, J., Marondedze, A., Masvaya, E.N., Mawodza, T., Nyawasha, R., Nyengerai, K., Tirivavi, R., Nyamugafata, P., Wuta, M., 2014. Influence of basin-based conservation

agriculture on selected soil quality parameters under smallholder farming in Zimbabwe. Soil Use Manag. 30, 550–559. https://doi.org/10.1111/sum.12149

Padarian, J., Minasny, B., McBratney, A.B., 2019. Using deep learning to predict soil properties from regional spectral data. Geoderma Reg. 16, e00198. https://doi.org/10.1016/j.geodrs.2018.e00198

Pedro, A.M.K., Ferreira, M.M.C., 2007. Simultaneously calibrating solids, sugars and acidity of tomato products using PLS2 and NIR spectroscopy. Anal. Chim. Acta 595, 221–227. https://doi.org/10.1016/j.aca.2007.03.036

Poeplau, C., Don, A., 2023. A simple soil organic carbon level metric beyond the organic carbon-to-clay ratio. Soil Use Manag. 39, 1057–1067. https://doi.org/10.1111/sum.12921

Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D., Pande, V., 2015. Massively Multitask Networks for Drug Discovery, in: International Conference on Machine Learning. https://doi.org/10.48550/arXiv.1502.02072

Shao, Y., He, Y., 2011. Nitrogen, phosphorus, and potassium prediction in soils, using infrared spectroscopy. Soil Res. 49, 166–172. https://doi.org/10.1071/SR10098

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., MacDonald, L.M., McLaughlin, M.J., 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Appl. Spectrosc. Rev. 49, 139–186. https://doi.org/10.1080/05704928.2013.811081

Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158, 46–54. https://doi.org/10.1016/j.geoderma.2009.12.025

Viscarra Rossel, R.A., Lark, R.M., 2009. Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. Eur. J. Soil Sci. 60, 453–464. https://doi.org/10.1111/j.1365-2389.2009.01121.x

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75. https://doi.org/10.1016/j.geoderma.2005.03.007

Vohland, M., Ludwig, M., Thiele-Bruhn, S., Ludwig, B., 2014. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. Geoderma 223–225, 88–96. https://doi.org/10.1016/j.geoderma.2014.01.013

Wadoux, A.M.J.-C., 2023. Interpretable spectroscopic modelling of soil with machine learning. Eur. J. Soil Sci. 74, 1–14. https://doi.org/10.1111/ejss.13370

Wadoux, A.M.J.-C., Malone, B., Minasny, B., Fajardo, M., McBratney, A.B., 2021. Soil Spectral Inference with R: Analysing Digital Soil Spectra Using the R Programming Environment, Springer International Publishing AG, Cham.

Wattel-Koekkoek, E.J.W., Van Genuchten, P.P.L., Buurman, P., Van Lagen, B., 2001. Amount and composition of clay-associated soil organic matter in a range of kaolinitic and smectitic soils. Geoderma 99, 27–49. https://doi.org/10.1016/S0016-7061(00)00062-8

Webster, R., Lark, R.M., 2013. Field Sampling for Environmental Science and Management. Routledge, New York. https://doi.org/10.4324/9780203128640

Wijewardane, N.K., Ge, Y., Wills, S., Libohova, Z., 2018. Predicting Physical and Chemical Properties of US Soils with a Mid-Infrared Reflectance Spectral Library. Soil Sci. Soc. Am. J. 82, 722–731. https://doi.org/10.2136/sssaj2017.10.0361

Wold, S., Albano, C., Dunn III, W.J., Esbensen, K., Hellberg, S., Johansson, E., Sjorstrom, S., 1983. Pattern recognition: Finding and using regularities in multi-variate data, in: Martens, H., Russwurm, Jr, H. (Eds.), Food Research and Data Analysis. Applied Science Publishers, London, pp. 147–188.

Zingore, S., Tittonell, P., Corbeels, M., van Wijk, M.T., Giller, K.E., 2011. Managing soil fertility diversity to enhance resource use efficiencies in smallholder farming systems: A case from Murewa District, Zimbabwe. Nutr. Cycl. Agroecosystems 90, 87–103. https://doi.org/10.1007/s10705-010-9414-0